Introduction to Digital Speech Processing Homework 1

FAQ

Q1. 如 2.0 投影片第二頁上方的圖， 第一個 state 對應了3個 o (o_1 o_2 o_3)， 第二個 state 對應了4個 o (o_4 o_5 o_6 o_7)， 第三個... 。 這些 state 每個分別對應到多少個 o，是不是必須先把 o_1 ~ o_T 都有了之後，再根據演算法得到最佳的一條路徑，才能推出每個 state 對應多少個 o (停留次數) ？

> 在 2.0 投影片第二頁的圖中，state 1 會有 3 個 observation，是因為頭兩次的 transition 都是 state 1 → state 1，這在 training 和 testing 的時候會因為 a_{11} 比 a_{12} , a_{13} ... 都大而顯現出來。換個角度說，在 2.0 投影片第二頁的圖中，state sequence 應該是 $q_1q_1q_1q_2q_2q_2q_2q_3$... 。這樣看的話就很清楚每個 state 是對應到一個 observation。因此演算法中 q 和 o 的下標都是從 1 到 T 。

Q2. train 出來的 model_01.txt ~ model_05.txt 該是什麼樣子?

> 就長得像 model_init.txt 一樣，其中 π 向量的總和需為 1， A 矩陣的每個 row sum 和 B 矩陣的每個 column sum 也都需是 1 。

Q3. 投影片裡第五張的 seq_model_01~05 都各自代表ㄅ的 HMM 嗎？ 就是說 seq_model_01 是ㄅ的 HMM，seq_model_02也是，這樣想對嗎？

> seq_model_01~05 可以想成是五個 phoneme 的 model，比方說知道以下的話只會出現ㄚ、ㄛ、ㄞ、ㄟ、ㄢ 五個音，於是就把ㄚ的 training data 合在一起 train 一個ㄚ的 model，ㄛ的 training data 合在一起 train 一個ㄛ的 model，...等等。於是有了五個 model 對應到五個 phoneme，在 testing 的時候就把 data 拿去在這五個 model 裡各求一個機率，如果是ㄟ的 model 機率最大就說答案是ㄟ。雖然現在我們沒有說 seq_model_01~05 是什麼東西，不過你可以想成 seq_model_01 是ㄚ的 training data，seq_model_02 是ㄛ的 training data 等等。

Q4. 如果我們有一筆 training data， state 數目是變動的，可能是 4 個 state 也可能是 5 個 state，那對於 4 個 state 來說，我如何用 trainging data 來算出這 4 個 state 上的 Observation 機率呢？ 改成定義 5 個 state 的話， Observation 的機率又該如何呢？ 而定義每個 state 的初始機率 A_{ij} 又是怎麼定義呢？

> 語音辨識中的 state 數目是使用者自己決定的，畢竟一個 phoneme 裡有很多 state 可說是我們因為語音連續性所做的假設。 因此究竟要訂定 4 個還是 5 個 state 並不是 data dependent，而是 user dependent。 使用 3~6 個 state 都是有看過的，甚至每個 model 都用不同數目的 state 理論上也是有可能的 (雖然實際上很少人這樣做)。 不過一但決定好，在 training/testing 流程中就不會再改變，因為演算法都是在假設 state 數目已知的情況下運作的。 在這個作業中只需要像 model_init.txt 裡一樣假設有 6 個 state 就好 (因此 A 是 6*6 矩陣，π 是 1*6 向量)。 至於初始機率，只要滿足適當的限制 (如 A 矩陣的每個 row sum 是 1) 即可，在 training 的 iteration 次數夠多的情況下應不至於對結果有太大的影響。

Q5. model_init.txt 的 observation 機率是否有問題呢? (為何不是根據 train_seq_0x.txt 去統計出 A 之機率多大... B 之機率多大去寫初始值...)

> 用 train_seq_0x.txt 去統計當做初始值當然也是可以的。 其實在語音辨識中，因為 B 的參數變成 Gaussian 的 mean 和 variance，它們的值無範圍限制難以隨便假設，此時有一種設定初始值的方法就是去算所有 observation 的 global 平均。 然而即使如此也無法分開 state 估計，每個 state 的初始值只能都設一樣， A_{ij} 的初始值也無法估計只能任意假設。 所以終究還是要跑 training algorithm。 而在跑過之後收斂到的結果雖然會跟初始值有關，但是

> 我們無法知道哪個初始值會產生較好的結果。 在作業中也可以嘗試不同的初始值，看看結果的差異。

Q6. 對 test_seq.txt 的資料要把每列的資料分別餵給 5 個 model 得到最大機率的就是答案，觀測值都固定是要 50 個嗎？ 要算這筆是某個 model 多大的機率，也是把這 50 個丟演算法得到？

> 在每一行有幾個字母就是幾個觀測值，建議不要寫死，寫成讓程式讀出觀測值的個數比較好。 觀測值的個數並不用固定，training 和 testing 可以用不同的個數，甚至 training 裡頭或 testing 裡頭也可以每筆資料有不同的個數。 以語音辨識為例，觀測值的個數相當於錄音的長度。 雖然也許可以硬性規定每筆聲音資料都錄一樣長，但是這相當麻煩，實際上不需要假設每筆資料都一樣。 在作業裡為了方便所以才會都是 50 個。

Q7. 關於 observation 機率的 adjust ，以 Σ (γ)分子的部份除以分母的部份如何區分？上面是寫 $o_t = v_k$，但是還是不太清楚？

> 對每個 state i 和 時間 t ，你的程式都會算出 $\gamma_t(i)$，而 update B 矩陣的分子部分，是要把不同 observation 的 t 的 $\gamma_t(i)$累積起來。 舉例來說如果 observation 是 AABCCBFFAEDD...，那麼 update $b_i(A)$ 的分子部分就是 $\gamma_1(i) + \gamma_2(i) + \gamma_9(i) + ...$，update $b_i(B)$ 的分子部分就是 $\gamma_3(i) + \gamma_6(i) + ...$ 等等。

Q8. 我想用 C 寫，但是不太懂 Makefile，如果要編譯成 .exe 檔要怎麼做？

> 寫 Makefile 的用意代表要在 Linux command line 下執行，不過這不代表一定要在自己的電腦灌 Linux，可以匯入助教設定好的虛擬機(username=root, password=ntudsp)，照著教學匯入即可。 Linux 下的基本指令操作可參考鳥哥，應該只需學會簡單的複製移動檔案指令就夠用了。 在 Linux 下並非以副檔名而是以權限作為能否當作執行檔的依據。 作業當中 Makefile 的作用就是從 C 的 source code 編譯出執行檔。
>
> 如果不想在虛擬機上做，也可以使用一種在 Windows 下模擬 Linux 環境的程式，叫做 Cygwin
>
> - 官方網站
> - 安裝教學: (1) | (2) | (3)
>
> > 在安裝選單中，請裝上：make、gcc、bash
>
> 或是也可以考慮 WSL (Windows Subsystem for Linux)，同樣也能在 Windows 下使用 Linux 環境。
>
> - 安裝教學

Q9. 編譯的時候 hmm.h 會有 warning 該怎麼辦?

> warning 是因為 hmm.h 裡面有使用 fscanf，但卻沒有理會他的回傳值。 這對於這次作業不會有影響，同學們可以不用理會它，或是用一個變數接收 fscanf 的回傳值，就不會產生 warning。

Q10. test_hmm.c 裡最前面那段 load_models 為甚麼要註解掉？

> 那段是介紹如何用 load_models 和 dump_models 這兩個函數，它們能一次讀取和印出 5 個名稱列在 modellist.txt 的 model，當然這 5 個 model 要都已經存在於資料夾中。你可以複製幾個 model 檔然後觀察他的效果。

Q11. train_seq_01~05.txt 裡面的資料有 10000 筆，是否每一行代表一次 iteration？

> 每一行代表一筆 sample，但是 training 時每次 update 都是把每一筆 training data 的 γ 和 ε 累加起來，相當於 4.0 投影片 update A, B 的公式，分子分母外面都多一層 Σ。 因此在投

> 影片 16 頁的 Σ 不只是對每個 t ，也是對每筆 sample 的 γ, ε 做加總，所以公式應修正如下：
>
> $$\pi_i = \frac{\sum_{n=1}^{N} \gamma_1^n(i)}{N}$$
>
> $$\overline{a}_{ij} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T-1} \epsilon_t^n(i,j)}{\sum_{n=1}^{N} \sum_{t=1}^{T-1} \gamma_t^n(i)}$$
>
> $$\bar{b}_j(k) = \text{Prob}\left[o_t = v_k \,|\, q_t = j\right] = \frac{\sum_{n=1}^{N} \sum_{t=1,\ o_t=v_k}^{T} \gamma_t^n(j)}{\sum_{n=1}^{N} \sum_{t=1}^{T} \gamma_t^n(j)}$$
>
> 上式中的 N 是 sample 的個數。

Q12. iteration 次數增加，accuracy 卻變低，是正常的嗎？

> 很正常，可以參考hw1投影片17頁的圖。

Q13. 請問一分鐘的時間限制是所有 model 的訓練時間合起來計時，還是每一個 model 都有一分鐘呢？

> 助教在使用你們的 training program 訓練多個 model 時，每個 model 各自都有一分鐘的訓練時間限制。

Q14. 請問 test_lbl.txt 檔是用來做什麼的，因為它好像不是程式輸入或輸出的檔案？

> TL;DR: 它是 test_seq.txt 的答案。
>
> 在 training 過程中，我們所使用的每個 train_seq_0x.txt 各自都是由單一 HMM 模型產出的 sequences； 而在 testing 的時候，我們所使用的 test_seq.txt 則是由多個 HMM 模型產出的 sequences 混合出來的。 所以才需要一份 test_lbl.txt (testing label)，讓大家在訓練和測試程式都寫完後，能夠將這份檔案和 testing program 的輸出比對一下，檢視自己的模型訓練的如何，或是看看測試程式有沒有寫對。

Q15. 請問評分時使用的 dataset 與提供給我們的那份一樣嗎？ 如果不同，請問評分用的 dataset 的 state 數量、observation 數量及 sequence 長度與提供給我們的 dataset 相同嗎？

> 評分時使用的 dataset 不是提供給同學們的那一組，但是它的 state 數量、observation 總數、sequence 長度以及 sequences 數量都與出作業時提供的那一組相同。